Available online at www.sciencedirect.com

SciVerse ScienceDirect

ELSEVIER

Journal of Approximation Theory 164 (2012) 1238–1261

Journal of Approximation Theory

www.elsevier.com/locate/jat

Full length article

Szegő's theorem for matrix orthogonal polynomials

Maxim Derevyagin[a], Olga Holtz[a,b,*], Sergey Khrushchev[c], Mikhail Tyaglov[a]

[a] Department of Mathematics MA 3-6, Technische Universität Berlin, Strasse des 17. Juni 136, D-10623 Berlin, Germany

[b] Department of Mathematics, University of California-Berkeley, 821 Evans Hall, Berkeley, CA, 94720, USA

[c] International School of Economics, Kazakh-British Technical University, Tole bi 59, 050000 Almaty, Kazakhstan

Received 21 December 2011; received in revised form 17 April 2012; accepted 3 May 2012
Available online 14 May 2012

Communicated by Leonid Golinskii

Abstract

We extend some classical theorems in the theory of orthogonal polynomials on the unit circle to the matrix case. In particular, we prove a matrix analogue of Szegő's theorem. As a by-product, we also obtain an elementary proof of the distance formula by Helson and Lowdenslager.

Keywords: Matrix orthogonal polynomials; Operator functions; Szegő's theory

1. Historical background and motivation

The classical work [16] of Szegő was the first to address the asymptotics of orthogonal polynomials on the unit circle $\mathbb{T}$ under the assumption that the entropy of the underlying measure σ is finite, i.e.,

$$\int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi} > -\infty.$$

* Corresponding author at: Department of Mathematics, University of California-Berkeley, 821 Evans Hall, Berkeley, CA, 94720, USA.
E-mail address: holtz@Math.Berkeley.EDU (O. Holtz).

0021-9045/$ - see front matter
doi:10.1016/j.jat.2012.05.003

Further aspects of Szegő's theory were developed by Geronimus, Verblunsky and others, which led to a number of other formulas, in various setups, involving the entropy such as the formula of Helson–Lowdenslager [9] for multivariate random processes (for a historical account, see [15, Section 1.1]).

Verblunsky [17, formulas (v) and (vi)] showed that, for any probability measure σ on the unit circle $\mathbb{T}$,

$$\lim_n \prod_{k=0}^{n}(1-|a_k|^2) = \exp\int_{\mathbb{T}} \log\sigma' \frac{d\theta}{2\pi}. \tag{1}$$

Here $\{a_k\}_{k\geq 0}$ is a sequence of points in the unit disc $\mathbb{D}$ called the *parameters* of σ [11, Section 8.3] and $\sigma' = 2\pi\, d\sigma/d\theta$ is the Lebesgue derivative of σ. The numbers $\{a_k\}_{k\geq 0}$ have different names depending on the area where they are considered. In the theory of orthogonal polynomials they are known as the *Szegő recurrence coefficients, Verblunsky parameters, Geronimus parameters,* in Schur's theory they are *Schur's parameters,* in inverse scattering problems they are *reflection coefficients;* see [15, Section 1.1].

In the matrix setting, σ is a Borel measure on $\mathbb{T}$ with values in the set $\mathcal{M}_\ell^+$ of all nonnegative definite matrices in $\mathcal{M}_\ell$, the set of all $\ell \times \ell$ matrices with complex entries. We denote by $\mathsf{P}_\ell(\mathbb{T})$ the set of all matrix-valued nonnegative measures σ on $\mathbb{T}$ that are normalized, i.e.,

$$\sigma(\mathbb{T}) = \mathbf{1}$$

to the unit matrix $\mathbf{1}$ in $\mathcal{M}_\ell^+$. We refer to $\mathsf{P}_\ell(\mathbb{T})$ as the class of *matrix probability measures.*

The matrix case is important in multivariate Time Series and Prediction Theory [8,9,13,14,18]. As far as we know, the first Szegő-type results on matrix-valued orthogonal polynomials were obtained by Delsarte et al. [5]; this line of research was continued by Aptekarev and Nikishin [2].

Recently, Damanik et al. set forth the following challenge [4].

Among the deepest and most elegant methods in OPUC are those of Khrushchev [125, 126, 101]. *We have not been able to extend them to MOPUC! We regard their extension as an important open question...*

Below we respond to this challenge, providing a full matrix-valued version of Szegő's theorem, yielding the previously known trace versions as corollaries of our matrix formula. Our method is a combination of the recent theory of matrix orthogonal polynomials presented in [4] and the approach to Szegő's theory developed in [12,11]. This combination allows us to avoid using factorization theory in matrix Hardy classes. Instead, we use only methods of Real Analysis and Matrix/Operator Theory.

Throughout the paper, we mostly follow the notation and terminology of [4].

2. Main results

In the matrix case, the parameters α_k are matrices in $\mathcal{M}_\ell$ with norms $\|\alpha_k\|$ not exceeding 1. Here $\|\alpha\|$ is the norm of the linear operator defined by the matrix α subordinate to the usual Euclidean vector norm (2-norm) on $\mathbb{C}^\ell$. This operator norm is also known as the *spectral* or the *Euclidean* norm. This norm is well known to equal the largest singular value of the matrix α; in particular, if α is self-adjoint, the norm $\|\alpha\|$ equals the spectral radius of α.

We denote by $\alpha^\dagger$ the Hermitian conjugate of $\alpha \in \mathcal{M}_\ell$. The symbol * is reserved for the Szegő dual, so we do not use it for the adjoint (see (4)).

We assume that the matrix

$$\int_{\mathbb{T}} p(e^{i\theta})^\dagger d\sigma(\theta) p(e^{i\theta}) \tag{2}$$

is positive (definite) for any polynomial p with coefficients in $\mathcal{M}_\ell$. Under this condition, the right (left) orthogonal matrix polynomials φ_n^R (φ_n^L) are uniquely determined by the standard Gram–Schmidt orthonormalization (more details are given in Section 6, also see [4]). It is important to note that the left orthogonal matrix polynomials are obtained with respect to the left quadratic 'form':

$$\int_{\mathbb{T}} p(e^{i\theta}) d\sigma(\theta) p(e^{i\theta})^\dagger. \tag{3}$$

Every $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ is uniquely determined by the sequence of its parameters $\{\alpha_k\}_{k\geq 0}$. These parameters are contractive matrices in $\mathcal{M}_\ell$. If σ is a matrix-valued measure with parameters $\{\alpha_k\}_{k\geq 0}$, then the parameters $\{\alpha_k^\dagger\}_{k\geq 0}$ correspond to the measure $\overline{\sigma}$ such that

$$\overline{\sigma}(E) = \sigma(\overline{E}), \quad \overline{E} = \{\overline{z} : z \in E\},$$

for any Borel set E, where $\overline{z}$ stands for the complex conjugate of a complex number z. We write $\varphi_n(z, \sigma)$ for the orthogonal polynomials if the dependence on σ is important.

For a matrix polynomial P_n of degree n, we define the *reversed* (or *Szegő dual)* polynomial P_n^* by

$$P_n^*(z) = z^n P_n(1/\overline{z})^\dagger. \tag{4}$$

The relationship between the left orthogonal polynomials φ_n^L and the right orthogonal polynomials φ_n^R is given by the formula

$$\varphi_n^L(e^{i\theta}, \overline{\sigma}) = \varphi_n^R(e^{-i\theta}, \sigma)^\dagger \tag{5}$$

(see Corollary 16). The nth left normalized orthogonal polynomial $\varphi_n^L(z, \overline{\sigma})$ depends on the parameters $\alpha_0^\dagger, \alpha_1^\dagger, \ldots, \alpha_{n-1}^\dagger$. Hence, the nth right polynomial $\varphi_n^R(z, \sigma)$ can be obtained from the left $\varphi_n^L(z, \sigma)$ by replacing each $\alpha_k^\dagger$ by α_k, replacing $z \in \mathbb{T}$ by $\overline{z}$ and applying the conjugation $\dagger$.

The main result of this paper is the following theorem.

Theorem. *Every matrix probability measure $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ satisfies the following matrix equality:*

$$\lim_{n\to+\infty} \int_0^{2\pi} \log([\varphi_n^{R,*}(e^{i\theta})^\dagger \varphi_n^{R,*}(e^{i\theta})]^{-1}) \frac{d\theta}{2\pi} = \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi}. \tag{6}$$

If the parameters α_k of σ form a family of commuting normal matrices, then (6) can be simplified to

$$\prod_{k=0}^{\infty} (\mathbf{1} - \alpha_k \alpha_k^\dagger) = \exp \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi}. \tag{7}$$

Remark. Alternatively, the commuting case reduces to the diagonal and hence to the scalar case.

Regardless of the normality or commutativity of $\{\alpha_k\}$, the following determinantal-trace version [5] follows from (6):

$$\prod_{k=0}^{\infty} \det\ (\mathbf{1} - \alpha_k \alpha_k^\dagger) = \exp \int_{\mathbb{T}} \operatorname{tr} \log \sigma' \frac{d\theta}{2\pi}. \tag{8}$$

The symmetry $z \mapsto \bar{z}$ keeps the Lebesgue measure on $\mathbb{T}$ invariant. Hence, combining (5) and a simple formula

$$\int_{\mathbb{T}} \log \overline{\sigma}' \frac{d\theta}{2\pi} = \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi},$$

we also obtain the left version of (6):

$$\lim_{n\to+\infty} \int_0^{2\pi} \log([\varphi_n^{L,*}(e^{i\theta})\varphi_n^{L,*}(e^{i\theta})^\dagger]^{-1}) \frac{d\theta}{2\pi} = \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi}. \tag{9}$$

3. Matrix preliminaries

Recall that we denote by $\mathcal{M}_\ell$ the ring of all $\ell \times \ell$ complex-valued matrices, its identity matrix by $\mathbf{1}$ and its zero matrix by $\mathbf{0}$. Along with the Euclidean norm $\|\cdot\|$ on $\mathcal{M}_\ell$, we also consider the trace norm $|\alpha|_1 = \operatorname{tr}(\alpha^\dagger\alpha)^{1/2}$ and the Hilbert–Schmidt norm $|\alpha|_2 = (\operatorname{tr}(\alpha^\dagger\alpha))^{1/2}$. It is easy to see that

$$\|\alpha\| \le |\alpha|_2 \le |\alpha|_1 \le \ell\|\alpha\|. \tag{10}$$

We say that a self-adjoint matrix $A(= A^\dagger) \in \mathcal{M}_\ell$ is *nonnegative* (*positive*) if the corresponding quadratic form $x \mapsto x^\dagger Ax$ is nonnegative definite (positive definite). We denote the class of all nonnegative self-adjoint $\ell \times \ell$ matrices by $\mathcal{M}_\ell^+$. The corresponding partial order is known as the *Loewner ordering* and is denoted by $\succ$: $A \succ B$ means that $A - B$ is positive, i.e., $A - B \succ \mathbf{0}$, and $A \succeq B$ means that $A - B \succeq \mathbf{0}$, or $A - B \in \mathcal{M}_\ell^+$.

Here is the first fact about the Loewner ordering that we will use later.

Lemma 1. *Let* $\mathbf{0} \preceq A_j \preceq B_j$ *for* $j = 1, \ldots k$. *Then* $\mathbf{0} \preceq A_1 + \cdots + A_k \preceq B_1 + \cdots + B_k$.

Proof. Evaluate and compare the quadratic forms of both sums. □

We will also need the following result connecting traces of self-adjoint matrices and their Loewner ordering.

Lemma 2. *Suppose* $A \succeq B$ *and* $\operatorname{tr}A = \operatorname{tr}B$. *Then* $A = B$.

Proof. By the linearity of traces, this is equivalent to the statement: suppose $A \succeq \mathbf{0}$ and $\operatorname{tr}A = 0$, then $A = \mathbf{0}$. The latter follows from the fact that $\operatorname{tr}A = \sum_{j=1}^{\ell} e_j^\dagger A e_j$, so if the trace of A is zero, the action of A on all standard unit vectors (hence on the entire space) must be trivial. □

Another fact about traces we will need is the following.

Lemma 3. *If* $A \succ \mathbf{0}$, *then* $\log\det(A) = \operatorname{tr}(\log A)$.

Proof. Without loss of generality, A is a diagonal matrix with positive diagonal elements, since the formula is invariant under unitary similarity. But then $\log A$ is the diagonal matrix whose elements are the logarithms of the diagonal elements of A. The conclusion of the Lemma is thus straightforward. $\square$

Corollary 4. *Let* $A_1, \ldots, A_n \succ \mathbf{0}$ *and let* $A := A_1 \cdots A_n \succ \mathbf{0}$. *Then*

$$\operatorname{tr}\log(A_1 \cdots A_n) = \log\det(A_1 \cdots A_n) = \sum_{k=1}^{n} \operatorname{tr}(\log A_k).$$

Proof. Apply Lemma 3 to the product $A = A_1 \cdots A_n$. $\square$

Finally, we will need the following interesting characterization of the determinant via the trace (also used by Helson and Lowdenslager in [9], also see, e.g., [10, Exercise 19, p. 486]).

Lemma 5. *Let* $\mathcal{A}$ *be the set of all matrices in* $\mathcal{M}_\ell$ *with determinant* 1. *Then every positive matrix* C *satisfies*

$$\inf_{A \in \mathcal{A}} \frac{1}{\ell} \operatorname{tr}(ACA^\dagger) = [\det(C)]^{1/\ell}. \tag{11}$$

Proof. Let U be a unitary matrix such that $C = UDU^\dagger$, where D is the diagonal matrix with eigenvalues $\lambda_1, \ldots \lambda_\ell$. Then $\lambda = \det(U) \in \mathbb{T}$. It follows that $ACA^\dagger$=$(A\overline{\lambda}U)D(A\overline{\lambda}U)^\dagger$, implying that we may assume without loss of generality that $C = D$. Then

$$\operatorname{tr}(ADA^\dagger) = \lambda_1\|a_1\|^2 + \lambda_2\|a_2\|^2 + \cdots + \lambda_\ell\|a_\ell\|^2,$$

where a_k denotes the kth column of A. By the arithmetic–geometric mean inequality,

$$\frac{\lambda_1\|a_1\|^2 + \lambda_2\|a_2\|^2 + \cdots + \lambda_\ell\|a_\ell\|^2}{\ell} \geq \sqrt[\ell]{\lambda_1 \cdots \lambda_\ell \|a_1\|^2 \cdots \|a_\ell\|^2}. \tag{12}$$

By Hadamard's inequality [10, Inequality 7.8.2],

$$\|a_1\| \cdots \|a_\ell\| \geq \det(A) = 1. \tag{13}$$

The equality in (12) occurs if and only if

$$\lambda_1\|a_1\|^2 = \cdots = \lambda_\ell\|a_\ell\|^2.$$

The equality in (13) occurs if and only if the columns a_k form an orthogonal system in $\mathbb{C}^\ell$. It follows that the equality in (11) is attained for the diagonal matrix A with $a\lambda_1^{-1/2}, \ldots, a\lambda_\ell^{-1/2}$ on the diagonal. Here a is chosen so as to make the determinant of A equal 1. $\square$

4. Matrix measures

A matrix-valued nonnegative measure μ on the unit circle $\mathbb{T}$ is a countably additive mapping of the Borel σ-algebra $\mathfrak{B}(\mathbb{T})$ on $\mathbb{T}$ into the set $\mathcal{M}_\ell^+$ of all nonnegative $\ell \times \ell$ matrices $\mu : B \mapsto \mu(B) \in \mathcal{M}_\ell^+$. It follows that for any $E \in \mathfrak{B}(\mathbb{T})$

$$\mathbf{0} \preceq \mu(E) \preceq \mu(E) + \mu(\mathbb{T} \setminus E) = \mu(\mathbb{T}).$$

Then $\nu(E) = \mu(\mathbb{T})^{-1/2}\mu(E)\mu(\mathbb{T})^{-1/2}$ is also a matrix-valued nonnegative measure which is called the normalization of μ. As before, we assume that μ is normalized: $\mu(\mathbb{T}) = \mathbf{1}$.

Recall that $\mathsf{P}_\ell(\mathbb{T})$ denotes the set of all matrix probability measures, i.e., the normalized matrix-valued nonnegative measures on $\mathbb{T}$. Let $\{e_j\}_{j=1}^{l}$ be the standard basis in $\mathbb{C}^\ell$. Then for every $E \in \mathfrak{B}(\mathbb{T})$ we obtain the matrix of $\mu(E)$

$$\mu(E) = \begin{pmatrix} \mu_{11}(E) & \mu_{12}(E) & \cdots & \mu_{1\ell}(E) \\ \mu_{21}(E) & \mu_{22}(E) & \cdots & \mu_{2\ell}(E) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{\ell 1}(E) & \mu_{\ell 2}(E) & \cdots & \mu_{\ell\ell}(E) \end{pmatrix}. \tag{14}$$

Since $|\alpha|_1 = \operatorname{tr}(\alpha)$ for every $\alpha \in \mathcal{M}_\ell^+$, we see that

$$|\mu_{ij}(E)| = |(\mu(E)e_j, e_i)| \le ||\mu(E)|| \le |\mu(E)|_1 = \operatorname{tr}(\mu(E)). \tag{15}$$

It follows that the entries $\mu_{ij}(E)$ of $\mu(E)$ are finite complex measures on $\mathbb{T}$ which are absolutely continuous with respect to $\operatorname{tr}(\mu)$. Thus any element μ of $\mathsf{P}_\ell(\mathbb{T})$ is nothing but a table of measures (14) subject to positivity conditions and domination by $\operatorname{tr}(\mu)$. We say that $\mu \in \mathsf{P}_\ell(\mathbb{T})$ is *absolutely continuous (discrete, singular)* if so is its trace measure. It follows that any $\mu \in \mathsf{P}_\ell(\mathbb{T})$ can be uniquely decomposed into the sum

$$\mu = \mu_a + \mu_d + \mu_s, \tag{16}$$

where μ_a is absolutely continuous with respect to the Lebesgue measure $d\theta/(2\pi)$, μ_d is the discrete part of μ and μ_s is its singular part. Indeed, taking the Hahn–Lebesgue decomposition of the trace measure, we can associate three matrix-valued measures with it. Namely, the entries of μ_a are the absolutely continuous parts of μ_{ij} with respect to $\operatorname{tr}(\mu)_a$, and similarly the entries of μ_d and μ_s for the discrete and singular parts, respectively. Since Borel supports of $\operatorname{tr}(\mu)_a$, $\operatorname{tr}(\mu)_d$, $\operatorname{tr}(\mu)_s$ can be chosen to be disjoint, the positivity of the corresponding matrices follows immediately. Moreover $d\mu = \mathbf{M}(\mu,\zeta)\operatorname{tr}(d\mu)$ where $\mathbf{M}(\mu,\zeta) \in \mathcal{M}_\ell^+$ for $\zeta \in \mathbb{T}$.

The measure μ_a can be found by Lebesgue's differentiation theorem:

$$\mu'(e^{i\theta}) = \lim_{\epsilon\to 0^+} \frac{\mu(I_\epsilon)}{2\epsilon} \quad \text{a.e. on } \mathbb{T}, \tag{17}$$

where I_ϵ denotes the arc of length 2ϵ on $\mathbb{T}$ centered at $e^{i\theta}$. Then

$$\mu_a(E) = \int_E \mu'(e^{i\theta})\frac{d\theta}{2\pi}.$$

We say that a sequence $\{\mu^{(n)}\}_{n\ge 0}$ in $\mathsf{P}_\ell(\mathbb{T})$ converges to $\mu \in \mathsf{P}_\ell(\mathbb{T})$ in the $*$-weak topology if

$$*{-}\lim_n \mu_{ij}^{(n)} = \mu_{ij} \tag{18}$$

for every pair of indices (i, j). For our class $\mathsf{P}_\ell(\mathbb{T})$, we need matrix analogues of two Helley's lemmas as they are stated in [11, Lemma 8.5, Theorem 8.6] for scalar measures.

Theorem 6. *If $*{-}\lim_n \mu^{(n)} = \mu$ in $\mathsf{P}_\ell(\mathbb{T})$, then*

$$\lim_n \mu^{(n)}(I) = \mu(I) \tag{19}$$

for any open arc I on $\mathbb{T}$ such that μ vanishes at the endpoints of I.

Proof. Let x be an arbitrary fixed column-vector in $\mathbb{C}^\ell$; as usual, $x^\dagger$ denotes its conjugate transpose (row-vector). Let $t \mapsto f(t)$ be a nonnegative continuous function with values in [0, 1] supported on an open arc I. Then

$$x^\dagger \mu^{(n)}(I)x = \int_I x^\dagger d\mu^{(n)}(t)x \geq \int_I f(t)x^\dagger d\mu^{(n)}(t)x. \tag{20}$$

By (18),

$$\begin{aligned}\lim_n \int_I f(t)x^\dagger d\mu^{(n)}(t)x &= \lim_n \int_I f(t) \sum_{i,j=1}^{\ell} \overline{x_i}\, d\mu_{ij}^{(n)} x_j \\ &= \sum_{i,j=1}^{\ell} \lim_n \int_I f(t)\overline{x_i}\, d\mu_{ij}^{(n)} x_j = \int_I f(t)x^\dagger d\mu(t)x.\end{aligned}$$

This observation and (20) imply

$$\begin{aligned}\limsup_n x^\dagger \mu^{(n)}(I)x &\geq \liminf_n x^\dagger \mu^{(n)}(I)x \geq \sup_f \int_I f(t)x^\dagger d\mu(t)x \\ &= \int_I x^\dagger d\mu(t)x = x^\dagger \mu(I)x.\end{aligned} \tag{21}$$

Similarly for the arc J complementary to the closure of I in $\mathbb{T}$ we have

$$\begin{aligned}\limsup_n x^\dagger \mu^{(n)}(J)x &\geq \liminf_n x^\dagger \mu^{(n)}(J)x \geq \sup_f \int_J f(t)x^\dagger d\mu(t)x \\ &= \int_J x^\dagger d\mu(t)x = x^\dagger \mu(J)x.\end{aligned} \tag{22}$$

Since μ vanishes at the endpoints of I and $\mu^{(n)} \in \mathsf{P}_\ell(\mathbb{T})$ we obtain that

$$\mu^{(n)}(I) + \mu^{(n)}(J) \preceq \mathbf{1}, \qquad \mu(I) + \mu(J) = \mathbf{1}. \tag{23}$$

Combining (21) and (22) with (23) we conclude that

$$\begin{aligned}x^\dagger x &= x^\dagger \mathbf{1} x \geq \limsup_n x^\dagger(\mu^{(n)}(I) + \mu^{(n)}(J))x \geq \liminf_n x^\dagger(\mu^{(n)}(I) + \mu^{(n)}(J))x \\ &\geq \int_I x^\dagger d\mu(t)x + \int_J x^\dagger d\mu(t)x = x^\dagger\mu(I)x + x^\dagger\mu(J)x = x^\dagger\mu(\mathbb{T})x = x^\dagger x.\end{aligned}$$

This is only possible if equalities hold in (21) and in (22). Since the vector x was arbitrary, this implies the conclusion of the theorem. □

Theorem 7. *Let $\{\mu^{(n)}\}_{n\geq 0}$ be a sequence in $\mathsf{P}_\ell(\mathbb{T})$ and $\mu \in \mathsf{P}_\ell(\mathbb{T})$. Then $*-\lim_n \mu^{(n)} = \mu$ if and only if $\lim_n \mu^{(n)}(I) = \mu(I)$ for any open arc I on $\mathbb{T}$ such that μ does not have point masses at the endpoints of I.*

Proof. One direction has been already proved in Theorem 6. Suppose now that $\lim_n \mu^{(n)}(I) = \mu(I)$ for any open arc I on $\mathbb{T}$ such that μ does not have point masses at the endpoints of I. Then for every $x \in \mathbb{C}^\ell$, $\|x\| = 1$, the sequence of usual probability measures $x^\dagger\mu^{(n)}x$ converges to $x^\dagger\mu x$ on any interval which does not have point masses of μ (and therefore of $x^\dagger\mu x$ since it is absolutely continuous with respect to μ). Then by Theorem 8.6 of [11] the sequence $x^\dagger \mu^{(n)} x$ converges to $x^\dagger \mu x$ in the $*$-weak topology. Now the polarization identity implies $*-\lim_n x^\dagger \mu^{(n)} y = x^\dagger \mu y$ for any pair of vectors (x, y). Setting $x := e_i$, $y := e_j$ for all pairs i, j, we obtain the weak limits for all entries of μ. $\square$

Theorem 8. *Suppose that $\nu_n = h_n d\theta/(2\pi)$ where h_n are matrix-valued functions on $\mathbb{T}$. Suppose that there is a positive constant C such that*

$$\int_{\mathbb{T}} \|h_n\|^2 \frac{d\theta}{2\pi} < C.$$

Then any $*$-weak limit point of $\{\nu_n\}$ is an absolutely continuous matrix-valued measure.

Proof. By norm equivalence (10) in $\mathcal{M}_\ell$, we can replace the operator norm of h_n by its Hilbert–Schmidt norm. For each matrix entry, the result of this theorem is standard, so it holds for the Hilbert–Schmidt (and hence the spectral) norm of the entire matrix as well. $\square$

Every $\mu \in \mathsf{P}_\ell(\mathbb{T})$ defines two positive definite quadratic forms on the two-sided module $C(\mathbb{T}, \mathcal{M}_\ell)$ over $\mathcal{M}_\ell$ of all continuous functions with values in $\mathcal{M}_\ell$. They correspond to the right and the left multiplication and are defined as matrix-valued 'inner products' by

$$\langle\langle f, g\rangle\rangle_R := \int f(x)^\dagger \, d\mu(x) g(x), \tag{24}$$

$$\langle\langle f, g\rangle\rangle_L := \int g(x) \, d\mu(x) f(x)^\dagger. \tag{25}$$

Let $\mathcal{P}$ denote the set of all polynomials in $z \in \mathbb{C}$ with coefficients from $\mathcal{M}_\ell$. For a nonnegative integer n, $\mathcal{P}_n$ will denote the set of polynomials in $\mathcal{P}$ of degree at most n. Note that, to generate an infinite sequence of orthogonal polynomials, μ must satisfy (2) for every nonzero polynomial p. This is equivalent to the condition that the non-negative Borel measure

$$\det(\mathbf{M}(\mu, \zeta))\mathrm{tr}(d\mu)$$

has infinite Borel support; see [18].

5. Analysis of operator functions

In this section we list some properties of the logarithm as an operator function. We start with the definitions of operator monotone, convex, and concave functions defined on the real half line $(0, \infty)$. Let $\mathcal{H}$ be an infinite-dimensional (separable) Hilbert space. Let $B_+(\mathcal{H})$ denote the set of all positive operators in $B(\mathcal{H})$. A continuous real function f on $(0, \infty)$ is said to be *operator monotone* (or, more precisely, *operator monotone increasing)* if $A \preceq B$ implies $f(A) \preceq f(B)$ for $A, B \in B_+(\mathcal{H})$, and *operator monotone decreasing* if $-f$ is operator monotone increasing, i.e., if $A \preceq B$ implies $f(A) \succeq f(B)$, where $f(A)$ and $f(B)$ are defined via functional calculus as usual. Also, f is said to be *operator convex* if $f(\lambda A + (1-\lambda)B) \preceq \lambda f(A) + (1-\lambda) f(B)$ for all $A, B \in B_+(\mathcal{H})$ and $\lambda \in (0, 1)$, and *operator concave* if $-f$ is operator convex (see also [3]).

One should not expect that the operator monotonicity and the operator convexity of f follow from the same properties of the scalar function f. For example, a power function t^α on $(0, \infty)$ is operator monotone if and only if $\alpha \in [0, 1]$, operator monotone decreasing if and only if $\alpha \in [-1, 0]$, and operator convex if and only if $\alpha \in [-1, 0] \cup [1, 2]$ (see, for instance, [3, Chapter V]). Moreover, the function $f(t) = \exp(t)$ is neither operator monotone nor operator convex on any (spectral) interval.

As is known, the operator monotone functions are generated by holomorphic functions that map the upper half plane into itself. Clearly, if one fixes a branch of the logarithm so that it is real on $(0, \infty)$ then the corresponding holomorphic function maps the upper half plane into itself.

Proposition 9. *The functions* $\log t$ *and* $-1/t$ *are operator monotone increasing on* $(0, \infty)$.

A detailed proof can be found in [3, Section V.4].

So, $1/t$ is operator monotone decreasing on $(0, \infty)$. Furthermore, it follows from [3, Exercise V.3.14] that the integration of an operator monotone decreasing function gives an operator concave function.

Proposition 10. *The function* $\log t$ *is operator concave on* $(0, \infty)$.

This statement can also be verified by means of [1, Theorem 3.1].

Now, we are in a position to formulate the matrix Jensen inequality for the logarithm. Namely, Proposition 10 and [7, Theorem 4.2] yield the following statement.

Proposition 11. *Let* $f : \mathbb{T} \to B_+(\mathcal{M}_\ell)$ *be a measurable function. Then the following inequality holds:*

$$\int_{\mathbb{T}} \log f(\theta) \frac{d\theta}{2\pi} \preceq \log \left(\int_{\mathbb{T}} f(\theta) \frac{d\theta}{2\pi} \right). \tag{26}$$

Besides monotonicity and convexity, we will also deal with operator continuity. Recall that a function f defined on $(0, \infty)$ is operator continuous if the relation $\|A_n - A\|_{\mathcal{H}} \to 0$ implies $\|f(A_n) - f(A)\|_{\mathcal{H}} \to 0$ for any $A, A_n \in B_+(\mathcal{H})$.

Proposition 12. *The function* $\log t$ *is operator continuous on* $(0, \infty)$.

Proof. Since $\log t$ can be extended to a holomorphic function on $\mathbb{C} \setminus (-\infty, 0]$, the statement follows directly from the Riesz–Dunford holomorphic functional calculus (see, e.g., [6]). $\square$

6. Matrix orthogonal polynomials on the unit circle

We begin by recalling some basic facts from [4] for the convenience of the reader. Let $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ be a matrix probability measure such that $\det(\mathbf{M}(\sigma, \zeta))\mathrm{tr}(d\sigma(\zeta))$ has infinite Borel support. We define right and left monic orthogonal matrix polynomials Φ_n^R, Φ_n^L by applying the Gram–Schmidt procedure in $C(\mathbb{T}, \mathcal{M}_\ell)$ with respect to the 'inner products' (24) and (25) to the sequence $\{\mathbf{1}, z\mathbf{1}, z^2\mathbf{1}, \ldots\}$. In other words, Φ_n^R is the unique matrix polynomial $z^n\mathbf{1}$ + *lower order terms* satisfying the orthogonality conditions

$$\mathbf{0} = \langle\langle z^k\mathbf{1}, \Phi_n^R \rangle\rangle_R = \int (z^k\mathbf{1})^\dagger d\sigma(x) \Phi_n^R, \quad k = 0, 1, \ldots, n-1. \tag{27}$$

Similarly, Φ_n^L is the unique matrix polynomial $z^n\mathbf{1}$ + *lower order terms* satisfying

$$\mathbf{0} = \langle\langle z^k\mathbf{1}, \Phi_n^L \rangle\rangle_L := \int \Phi_n^L(z)\, d\sigma(z)(z^k\mathbf{1})^\dagger, \quad k = 0, 1, \ldots, n-1. \tag{28}$$

The normalized orthogonal matrix polynomials are defined by

$$\varphi_0^L = \varphi_0^R = \mathbf{1}, \qquad \varphi_n^L = \kappa_n^L \Phi_n^L \quad \text{and} \quad \varphi_n^R = \Phi_n^R \kappa_n^R \tag{29}$$

where the κ's are defined according to the normalization conditions

$$\langle\langle \varphi_n^R, \varphi_m^R \rangle\rangle_R = \delta_{nm}\mathbf{1}, \qquad \langle\langle \varphi_n^L, \varphi_m^L \rangle\rangle_L = \delta_{nm}\mathbf{1}, \tag{30}$$

along with the following positivity conditions:

$$\kappa_{n+1}^L(\kappa_n^L)^{-1} \succ \mathbf{0} \quad \text{and} \quad (\kappa_n^R)^{-1}\kappa_{n+1}^R \succ \mathbf{0}. \tag{31}$$

Note that the κ_n^L are determined by the normalization condition up to multiplication on the left by unitary matrices. It can be shown that these unitaries can always be uniquely chosen so as to satisfy (31); see [4].

Now define

$$\rho_n^L := \kappa_n^L(\kappa_{n+1}^L)^{-1} \quad \text{and} \quad \rho_n^R := (\kappa_{n+1}^R)^{-1}\kappa_n^R.$$

Being inverses of positives matrices, ρ_n^L and ρ_n^R are positive definite as well. In particular, we have that

$$\kappa_n^L = (\rho_0^L \cdots \rho_{n-1}^L)^{-1} \quad \text{and} \quad \kappa_n^R = (\rho_{n-1}^R \cdots \rho_0^R)^{-1}. \tag{32}$$

In the matrix case as well as in the scalar case we have the Szegő recursion. Before stating it, we recall that, for a matrix polynomial P_n of degree n, we define the reversed polynomial P_n^* by (4): $P_n^*(z) = z^n P_n(1/\bar{z})^\dagger$.

Theorem 13 (*[4]*). *There is a sequence of contractive matrices α_n in $\mathcal{M}_\ell$ such that*

$$z\varphi_n^L - \rho_n^L\varphi_{n+1}^L = \alpha_n^\dagger\varphi_n^{R,*}, \tag{33}$$

$$z\varphi_n^R - \varphi_{n+1}^R\rho_n^R = \varphi_n^{L,*}\alpha_n^\dagger, \tag{34}$$

where ρ_n^L and ρ_n^R are defined as follows

$$\rho_n^L = (\mathbf{1} - \alpha_n^\dagger\alpha_n)^{1/2}, \qquad \rho_n^R = (\mathbf{1} - \alpha_n\alpha_n^\dagger)^{1/2}. \tag{35}$$

Setting $z = 0$ in (33) and using (29), we derive the following formulas for the parameters:

$$\alpha_n = -(\kappa_n^R)^{-1}\Phi_{n+1}^L(0)^\dagger(\kappa_n^L)^\dagger = -(\kappa_n^R)^\dagger\Phi_{n+1}^R(0)^\dagger(\kappa_n^L)^{-1}. \tag{36}$$

Alternatively, one can also set $z = 0$ in formulas (3.11) of [4].

Lemma 14. *The left and right monic orthogonal polynomials of $\overline{\sigma}$ and σ are related by*

$$\Phi_n^L(e^{i\theta}, \overline{\sigma}) = \Phi_n^R(e^{-i\theta}, \sigma)^\dagger. \tag{37}$$

Proof. For $k < n$ we have, by (27),

$$\begin{aligned}
\mathbf{0} &= \left(\int (z^k\mathbf{1})^\dagger\, d\sigma(z)\,\Phi_n^R(z,\sigma)\right)^\dagger = \int (\Phi_n^R(z,\sigma))^\dagger\, d\sigma(z)(z^k\mathbf{1}) \\
&= \int (\Phi_n^R(\bar{z},\sigma))^\dagger\, d\sigma(\bar{z})(\bar{z}^k\mathbf{1}) \\
&= \int (\Phi_n^R(\bar{z},\sigma))^\dagger\, d\overline{\sigma}(z)(z^k\mathbf{1})^\dagger = \int \Phi_n^L(z,\overline{\sigma})\, d\overline{\sigma}(z)(z^k\mathbf{1})^\dagger,
\end{aligned}$$

which implies (37) by (28). □

Proposition 15. *If $\{\alpha_k\}_{k\geq 0}$ are the parameters of σ, then $\{\alpha_k^\dagger\}_{k\geq 0}$ are the parameters of $\overline{\sigma}$.*

Proof. By (37), the matrix coefficients of the polynomial $\Phi_n^L(e^{i\theta}, \overline{\sigma})$ are the matrices adjoint to the coefficients of the polynomial $\Phi_n^R(e^{i\theta}, \sigma)$. In particular,

$$\Phi_{n+1}^L(0, \overline{\sigma}) = \Phi_{n+1}^R(0, \sigma)^\dagger. \tag{38}$$

Since $\kappa_0^R = \kappa_0^L = \mathbf{1}$, we see that

$$\alpha_0(\overline{\sigma}) = -\Phi_1^L(0, \overline{\sigma})^\dagger = -\Phi_1^R(0, \sigma) = \alpha_0(\sigma)^\dagger.$$

Suppose that we already proved that $\alpha_k(\overline{\sigma}) = \alpha_k(\sigma)^\dagger$ for $k < n$. Then, by the induction hypothesis and by (32),

$$\kappa_n^R(\overline{\sigma}) = \kappa_n^L(\sigma)^\dagger, \qquad \kappa_n^L(\overline{\sigma}) = \kappa_n^R(\sigma)^\dagger. \tag{39}$$

It follows that

$$\begin{aligned}\alpha_n(\overline{\sigma}) &= -(\kappa_n^R(\overline{\sigma}))^{-1}\Phi_{n+1}^L(0, \overline{\sigma})^\dagger(\kappa_n^L(\overline{\sigma}))^\dagger = -(\kappa_n^L(\sigma)^\dagger)^{-1}\Phi_{n+1}^R(0, \sigma)\kappa_n^R(\sigma)\\ &= (-\kappa_n^R(\sigma)^\dagger\Phi_{n+1}^R(0, \sigma)^\dagger(\kappa_n^L(\sigma))^{-1})^\dagger = \alpha_n(\sigma)^\dagger;\end{aligned}$$

see (36). □

Corollary 16. *The left and right orthogonal polynomials are related by formula (5).*

Proof. We have

$$\varphi_n^L(e^{i\theta}, \overline{\sigma}) = \kappa_n^L(\overline{\sigma})\Phi_n^L(e^{i\theta}, \overline{\sigma}) = \kappa_n^R(\sigma)^\dagger\Phi_n^R(e^{-i\theta}, \sigma)^\dagger = \varphi_n^R(e^{-i\theta}, \sigma)^\dagger. \quad \square$$

We next recall the notion of Bernstein–Szegő approximation. We begin with a list of properties of matrix orthogonal polynomials.

Theorem 17 (*[4, Theorem 3.9]*). *The polynomials φ^L, φ^R satisfy the following conditions.*

(i) *For $z \in \mathbb{T}$, all of $\varphi_n^{R,*}(z)$, $\varphi_n^{L,*}(z)$, $\varphi_n^R(z)$, $\varphi_n^L(z)$ are invertible.*
(ii) *For $z \in \mathbb{D}$, $\varphi_n^{R,*}(z)$ and $\varphi_n^{L,*}(z)$ are invertible.*
(iii) *For any $z \in \mathbb{T}$,*

$$\varphi_n^R(z)\varphi_n^R(z)^\dagger = \varphi_n^L(z)^\dagger\varphi_n^L(z). \tag{40}$$

Given a finite sequence $\{\alpha_j\}_{j=0}^{n-1}$ of contractive matrices, we can always use the Szegő recursion to define the polynomials φ_j^R, φ_j^L for $j = 0, 1, \ldots, n$. Analogously to the scalar case, let us define a measure $d\mu_n$ on $\mathbb{T}$ by

$$d\mu_n(\theta) = [\varphi_n^R(e^{i\theta})\varphi_n^R(e^{i\theta})^\dagger]^{-1}\frac{d\theta}{2\pi}. \tag{41}$$

In view of (40), we also see that

$$d\mu_n(\theta) = [\varphi_n^L(e^{i\theta})^\dagger\varphi_n^L(e^{i\theta})]^{-1}\frac{d\theta}{2\pi}. \tag{42}$$

Also, directly from the definition of the right orthogonal polynomials, we have

$$d\mu_n(\theta) = [\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1}\frac{d\theta}{2\pi}. \tag{43}$$

The measure $d\mu_n$ in (43) is called the *right Bernstein–Szegő approximation* to σ. The right-hand side of (43) can be also rewritten in yet another way as (cf. the discussion preceding Proposition 19)

$$d\mu_n(\theta) = [\varphi_n^{L,*}(e^{i\theta})\varphi_n^{L,*}(e^{i\theta})^\dagger]^{-1}\frac{d\theta}{2\pi}, \tag{44}$$

thus providing also the *left Bernstein–Szegő approximation* to σ. In other words, the right and the left Bernstein–Szegő approximations are the same, making the distinction between "right" and "left" approximations unnecessary from this point on. We now formulate the main result of this section.

Theorem 18 (*[4]*). *The matrix-valued measure $d\mu_n$ is normalized and its right matrix orthogonal polynomials for $j = 0, \dots, n$ are $\{\varphi_j^R\}_{j=0}^n$. The parameters of $d\mu_n$ are*

$$\alpha_j(d\mu_n) = \begin{cases}\alpha_j, & j \le n,\\ \mathbf{0}, & j \ge n+1.\end{cases} \tag{45}$$

Moreover, $*-\lim_{n\to\infty} d\mu_n = d\sigma$.

Following [4], we associate the matrix

$$A^L(\alpha, z) = \begin{pmatrix} z(\rho^L)^{-1} & -(\rho^L)^{-1}\alpha^\dagger \\ -z(\rho^R)^{-1}\alpha & (\rho^R)^{-1}\end{pmatrix} \tag{46}$$

to a given matrix parameter α. Then

$$\begin{pmatrix}\varphi_n^L\\ \varphi_n^{R,*}\end{pmatrix} = A^L(\alpha_{n-1}, z)\cdots A^L(\alpha_0, z)\begin{pmatrix}\mathbf{1}\\ \mathbf{1}\end{pmatrix}. \tag{47}$$

Applying the adjoint $\dagger$ to both sides and taking the product over α_j for $j = 0, \dots, n-1$, we obtain

$$\varphi_n^{L\dagger}\varphi_n^L + \varphi_n^{R,*\dagger}\varphi_n^{R,*} = (\mathbf{1}\ \mathbf{1})A^L(\alpha_0, z)^\dagger\cdots A^L(\alpha_{n-1}, z)^\dagger A^L(\alpha_{n-1}, z)\cdots A^L(\alpha_0, z)\begin{pmatrix}\mathbf{1}\\ \mathbf{1}\end{pmatrix}.$$

Note that (42) and (43) imply that the equality

$$\varphi_n^{L\dagger}\varphi_n^L = \varphi_n^{R,*\dagger}\varphi_n^{R,*} \tag{48}$$

holds on the circle $\mathbb{T}$, implying that

$$\varphi_n^{R,*\dagger}\varphi_n^{R,*} = \frac{1}{2}(\mathbf{1}\ \mathbf{1})A^L(\alpha_0, z)^\dagger\cdots A^L(\alpha_{n-1}, z)^\dagger A^L(\alpha_{n-1}, z)\cdots A^L(\alpha_0, z)\begin{pmatrix}\mathbf{1}\\ \mathbf{1}\end{pmatrix}. \tag{49}$$

Using the fact $\rho^R\alpha = \alpha\rho^L$, the matrix in (46) can be factored as follows:

$$\begin{aligned} A^L(\alpha, z) &= \begin{pmatrix}(\rho^L)^{-1} & \mathbf{0}\\ \mathbf{0} & (\rho^R)^{-1}\end{pmatrix}\begin{pmatrix} z\mathbf{1} & -\alpha^\dagger\\ -z\alpha & \mathbf{1}\end{pmatrix}\\ &= \begin{pmatrix} z\mathbf{1} & -\alpha^\dagger\\ -z\alpha & \mathbf{1}\end{pmatrix}\begin{pmatrix}(\rho^L)^{-1} & \mathbf{0}\\ \mathbf{0} & (\rho^R)^{-1}\end{pmatrix}. \end{aligned} \tag{50}$$

To factor the non-diagonal matrix in (50), we apply Schur's factorization

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ CA^{-1} & \mathbf{1} \end{pmatrix} \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} \mathbf{1} & A^{-1}B \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \tag{51}$$

with

$$\begin{aligned} A &= z\mathbf{1} \qquad B = -\alpha^\dagger \\ C &= -z\alpha \qquad D = \mathbf{1}. \end{aligned}$$

Then

$$A^L(\alpha, z) = \begin{pmatrix} (\rho^L)^{-1} & \mathbf{0} \\ \mathbf{0} & (\rho^R)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ -\alpha & \mathbf{1} \end{pmatrix} \begin{pmatrix} z\mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} - \alpha\alpha^\dagger \end{pmatrix} \begin{pmatrix} \mathbf{1} & -(z)^{-1}\alpha^\dagger \\ \mathbf{0} & \mathbf{1} \end{pmatrix}. \tag{52}$$

Since $\det \rho^L = \det \rho^R$, we conclude that

$$\det A^L(\alpha, z) = z^\ell. \tag{53}$$

It is easy to check that

$$\varphi_1^L(z) = (\rho_0^L)^{-1}(z - \alpha_0^\dagger), \qquad \varphi_1^R(z) = (z - \alpha_0^\dagger)(\rho_0^R)^{-1}.$$

Further, forming the Szegő dual, we obtain

$$\varphi_1^{L,*}(z) = (\mathbf{1} - z\alpha_0)(\rho_0^L)^{-1}, \qquad \varphi_1^{R,*}(z) = (\rho_0^R)^{-1}(\mathbf{1} - z\alpha_0).$$

After pertinent multiplications, this produces

$$\begin{aligned} \varphi_1^{R,*}(z)^\dagger \varphi_1^{R,*}(z) &= (\mathbf{1} - \bar{z}\alpha_0^\dagger)(\rho_0^R)^{-2}(\mathbf{1} - z\alpha_0), \\ \varphi_1^{L,*}(z)\varphi_1^{L,*}(z)^\dagger &= (\mathbf{1} - z\alpha_0)(\rho_0^L)^{-2}(\mathbf{1} - \bar{z}\alpha_0^\dagger). \end{aligned}$$

To compare these expressions for $\varphi_1^{R,*}(z)^\dagger \varphi_1^{R,*}(z)$ and $\varphi_1^{L,*}(z)\varphi_1^{L,*}(z)^\dagger$, we need the following observation.

Proposition 19. *For every $\alpha \in \mathcal{M}_\ell$ and $z \in \mathbb{T}$*

$$(\mathbf{1} - \bar{z}\alpha^\dagger)(\mathbf{1} - \alpha\alpha^\dagger)^{-1}(\mathbf{1} - z\alpha) = [(\mathbf{1} - \bar{z}\alpha^\dagger)^{-1} + (\mathbf{1} - z\alpha)^{-1} - \mathbf{1}]^{-1}.$$

Proof. Consider the matrix polynomial

$$p(z) = (\mathbf{1} - \bar{z}\alpha^\dagger)(\mathbf{1} - \alpha\alpha^\dagger)^{-1}(\mathbf{1} - z\alpha). \tag{54}$$

Since $z\bar{z} = 1$, we have

$$\mathbf{1} - \alpha\alpha^\dagger = \mathbf{1} - z\bar{z}\alpha\alpha^\dagger = (\mathbf{1} - z\alpha)(\mathbf{1} + \bar{z}\alpha^\dagger) + z\alpha - \bar{z}\alpha^\dagger. \tag{55}$$

Similarly,

$$\mathbf{1} - \alpha\alpha^\dagger = \mathbf{1} - z\bar{z}\alpha\alpha^\dagger = (\mathbf{1} + z\alpha)(\mathbf{1} - \bar{z}\alpha^\dagger) - z\alpha + \bar{z}\alpha^\dagger. \tag{56}$$

The sum of the expressions (55) and (56) yields

$$\begin{aligned} 2(\mathbf{1} - \alpha\alpha^\dagger) &= (\mathbf{1} + z\alpha)(\mathbf{1} - \bar{z}\alpha^\dagger) + (\mathbf{1} - z\alpha)(\mathbf{1} + \bar{z}\alpha^\dagger) \\ &= (\mathbf{2} - (\mathbf{1} - z\alpha))(\mathbf{1} - \bar{z}\alpha^\dagger) + (\mathbf{1} - z\alpha)(\mathbf{2} - (\mathbf{1} - \bar{z}\alpha^\dagger)). \end{aligned} \tag{57}$$

Let us denote $\mathbf{1} - z\alpha$ by B for brevity. From (54) and (57) we obtain

$$\begin{aligned} p(z) &= B^\dagger(\mathbf{1} - \alpha\alpha^\dagger)^{-1}B = 2B^\dagger[B(\mathbf{2} - B^\dagger) + (\mathbf{2} - B)B^\dagger]^{-1}B \\ &= 2[(\mathbf{2} - B^\dagger)B^{-\dagger} + B^{-1}(\mathbf{2} - B)]^{-1} = 2[2B^{-\dagger} - \mathbf{2} + 2B^{-1}]^{-1} \\ &= [(\mathbf{1} - \bar{z}\alpha^\dagger)^{-1} + (\mathbf{1} - z\alpha)^{-1} - \mathbf{1}]^{-1}. \quad \square \end{aligned}$$

As a corollary to Proposition 19, we obtain that $\varphi_1^{R,*}(z)^\dagger\varphi_1^{R,*}(z)$ and $\varphi_1^{L,*}(z)\varphi_1^{L,*}(z)^\dagger$ coincide.

7. The Bernstein–Szegő approximation

In this section we obtain a formula for the Bernstein–Szegő approximation of a matrix probability measure.

Lemma 20. *Let β_n be the matrix defined by*

$$\beta_n = \exp\int_0^{2\pi}\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi}. \tag{58}$$

Then $\log\beta_n$ *is self-adjoint and nonpositive.*

Proof. Since $[\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1}$ is a positive matrix for every θ its logarithm is self-adjoint as well as the integral $\log\beta_n$ of the logarithm of this matrix-valued function. By Proposition 11,

$$\begin{aligned} \log\beta_n &= \int_0^{2\pi}\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi} \\ &\preceq \log\left(\int_0^{2\pi}[\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1}\frac{d\theta}{2\pi}\right) \\ &= \log\mu_n(\mathbb{T}) = \mathbf{0} \end{aligned}$$

since μ_n is normalized: $\mu_n(\mathbb{T}) = \mathbf{1}$. $\square$

The standard operator calculus and Lemma 20 imply that the matrix $\beta_n = \exp\log\beta_n$ is self-adjoint and satisfies

$$\mathbf{0} \prec \beta_n \preceq \mathbf{1}. \tag{59}$$

Lemma 21. *For β_n, we have*

$$\log\det\beta_n = \operatorname{tr}\log\beta_n = \log\prod_{k=0}^{n-1}\det(\mathbf{1} - \alpha_k\alpha_k^\dagger) = \sum_{k=0}^{n-1}\operatorname{tr}\log(\mathbf{1} - \alpha_k\alpha_k^\dagger). \tag{60}$$

Proof. Applying elementary transformations and formula (35), we obtain

$$\begin{aligned} \operatorname{tr}(\log\beta_n) &= \int_0^{2\pi}\operatorname{tr}\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi} \\ &= \int_0^{2\pi}\log\det([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi} \end{aligned}$$

$$
\begin{aligned}
&= \int_0^{2\pi} \log\det([\varphi_n^{R,*}(e^{i\theta})^\dagger]^{-1})\frac{d\theta}{2\pi} + \int_0^{2\pi} \log\det([\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi} \\
&= 2\mathbf{Re}\int_0^{2\pi} \log\det([\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi} \\
&= 2\mathbf{Re}\log\det([\varphi_n^{R,*}(0)]^{-1}) = 2\log\det(\rho_0^R\cdots\rho_{n-1}^R) \\
&= \log\prod_{k=0}^{n-1}\det(\mathbf{1}-\alpha_k\alpha_k^\dagger)
\end{aligned}
$$

since the function $z \mapsto \log\det([\varphi_n^{R,*}(z)]^{-1})$ is analytic in the closed unit disc. $\square$

Remark. In general, $\log(AB)$ cannot be written as $\log A + \log B$ if A and B are matrices. So, the integral in Lemma 20 cannot be evaluated by the mean value theorem. In other words, the function $\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1})$ is in general not a restriction of a harmonic function to the unit circle. Our next lemma addresses the easy case when the logarithm in question does split.

If $\{\alpha_1,\ldots,\alpha_{n-1}\}$ is a commuting family of normal matrices, then the self-adjoint matrix β_n can be evaluated explicitly as follows.

Lemma 22. *Let $\{\alpha_1,\ldots,\alpha_{n-1}\}$ be a commuting family of normal matrices. Then*

$$\beta_n = \prod_{k=0}^{n-1}(\mathbf{1}-\alpha_k\alpha_k^\dagger).$$

Proof. The proof follows the proof of Lemma 21 since in this case $\varphi_n^{R,*}(z)$ is a normal matrix for any value of z. $\square$

8. The matrix Szegő theorem

Definition 23. A matrix probability measure $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ is said to be a Szegö measure if

$$\int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi} > -\infty. \tag{61}$$

Theorem 24. *For any matrix probability measure $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ and any $n \in \mathbb{N}$,*

$$\int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi} \le \operatorname{tr}\log\beta_n = \log\prod_{k=0}^{n-1}\det(1-\alpha_k^\dagger\alpha_k). \tag{62}$$

Proof. If $\int_{\mathbb{T}}\operatorname{tr}\log\sigma'\frac{d\theta}{2\pi} = -\infty$, the conclusion of the theorem holds trivially, so assume that σ is a Szegő measure, i.e., the corresponding integral is not $-\infty$. Jensen's matrix inequality from Proposition 11 implies

$$
\begin{aligned}
&\int_{\mathbb{T}} \log(\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger\sigma'\varphi_n^R(e^{i\theta})\beta_n^{1/2}])\frac{d\theta}{2\pi} \\
&\quad\preceq \log\left(\int_{\mathbb{T}}\beta_n^{1/2}\left[\varphi_n^R(e^{i\theta})^\dagger\sigma'\frac{d\theta}{2\pi}\varphi_n^R(e^{i\theta})\right]\beta_n^{1/2}\right).
\end{aligned} \tag{63}
$$

Now we replace σ' by σ, which is larger in the Loewner ordering, according to (16). Since σ is absolutely continuous with respect to $\mathrm{tr}(\sigma)$, there exist two disjoint Borel sets E and F and a Borel matrix function $x \mapsto \mathtt{M}(x)$ such that

$$d\sigma = \mathtt{M}\chi_E \mathrm{tr}(\sigma_a) + \mathtt{M}\chi_F \mathrm{tr}(d\sigma_d + d\sigma_s)$$

where E is a Borel support of $\mathrm{tr}(d\sigma_a)$, F is a Borel support of $\mathrm{tr}(d\sigma_d + d\sigma_s)$ and χ_E, χ_F are the indicators of E and F correspondingly. Notice that

$$\begin{aligned} &\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \mathtt{M}\chi_E \mathrm{tr}(\sigma_a)\varphi_n^R(e^{i\theta})]\beta_n^{1/2} = \beta_n^{1/2}\left[\varphi_n^R(e^{i\theta})^\dagger \sigma' \frac{d\theta}{2\pi}\varphi_n^R(e^{i\theta})\right]\beta_n^{1/2}; \\ &\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \mathtt{M}\chi_F \mathrm{tr}(d\sigma_d + d\sigma_s)\varphi_n^R(e^{i\theta})]\beta_n^{1/2} \\ &\quad = \beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger (d\sigma_d + d\sigma_s)\varphi_n^R(e^{i\theta})]\beta_n^{1/2}. \end{aligned} \tag{64}$$

Combining (64) with the result of Lemma 1, we obtain

$$\begin{aligned} &\int_{\mathbb{T}} \beta_n^{1/2}\left[\varphi_n^R(e^{i\theta})^\dagger \sigma' \frac{d\theta}{2\pi}\varphi_n^R(e^{i\theta})\right]\beta_n^{1/2} = \int_{\mathbb{T}} \beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \mathtt{M}\chi_E \mathrm{tr}(\sigma_a)\varphi_n^R(e^{i\theta})]\beta_n^{1/2} \\ &\quad \preceq \int_{\mathbb{T}} \beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \mathtt{M}\chi_E \mathrm{tr}(\sigma_a)\varphi_n^R(e^{i\theta})]\beta_n^{1/2} \\ &\qquad + \int_{\mathbb{T}} \beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \mathtt{M}\chi_F \mathrm{tr}(d\sigma_d + d\sigma_s)\varphi_n^R(e^{i\theta})]\beta_n^{1/2} \\ &\quad = \int_{\mathbb{T}} \beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger d\sigma \varphi_n^R(e^{i\theta})]\beta_n^{1/2}. \end{aligned}$$

By (63) and by the operator monotonicity of the logarithm from Proposition 9, we obtain

$$\begin{aligned} &\int_{\mathbb{T}} \log(\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \sigma' \varphi_n^R(e^{i\theta})]\beta_n^{1/2})\frac{d\theta}{2\pi} \\ &\quad \preceq \log\left(\int_{\mathbb{T}} \beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger d\sigma \varphi_n^R(e^{i\theta})]\beta_n^{1/2}\right) \\ &\quad = \log\left(\beta_n^{1/2}\int_{\mathbb{T}} [\varphi_n^R(e^{i\theta})^\dagger d\sigma \varphi_n^R(e^{i\theta})]\beta_n^{1/2}\right) = \log(\beta_n^{1/2}\mathbf{1}\beta_n^{1/2}) = \log \beta_n, \end{aligned} \tag{65}$$

in view of the orthonormality of the polynomials φ_n^R. Next, we have

$$\begin{aligned} &\mathrm{tr}\log(\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \sigma' \varphi_n^R(e^{i\theta})]\beta_n^{1/2}) = \log\det(\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger \sigma' \varphi_n^R(e^{i\theta})]\beta_n^{1/2}) \\ &\quad = \log[\det(\beta_n)\det([\varphi_n^R(e^{i\theta})^\dagger \varphi_n^R(e^{i\theta})])\det(\sigma')] \\ &\quad = \log\det(\beta_n) + \log\det([\varphi_n^R(e^{i\theta})^\dagger \varphi_n^R(e^{i\theta})]) + \log\det(\sigma') \\ &\quad = \mathrm{tr}\log\beta_n + \mathrm{tr}\log[\varphi_n^R(e^{i\theta})^\dagger \varphi_n^R(e^{i\theta})] + \mathrm{tr}\log\sigma'. \end{aligned}$$

Taking the trace in (65), integrating over $\mathbb{T}$ and taking into account (58), we arrive at

$$\begin{aligned}\operatorname{tr}\log\beta_n &\geq \int_{\mathbb{T}} \operatorname{tr}\log(\beta_n^{1/2}[\varphi_n^R(e^{i\theta})^\dagger\sigma'\varphi_n^R(e^{i\theta})]\beta_n^{1/2})\frac{d\theta}{2\pi}\\ &= \operatorname{tr}\log\beta_n + \int_{\mathbb{T}} \operatorname{tr}\log([\varphi_n^R(e^{i\theta})^\dagger\varphi_n^R(e^{i\theta})])\frac{d\theta}{2\pi} + \int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi}\\ &= \operatorname{tr}\log\beta_n + \int_{\mathbb{T}} \operatorname{tr}\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})])\frac{d\theta}{2\pi} + \int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi}\\ &= \int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi}.\end{aligned}$$

The second-to-last equality is due to the fact that $(\varphi_n^R)^\dagger\varphi_n^R = \varphi_n^{R,*}(\varphi_n^{R,*})^\dagger$ on the unit circle, and hence

$$\begin{aligned}\operatorname{tr}\log([\varphi_n^R(e^{i\theta})^\dagger\varphi_n^R(e^{i\theta})]) &= \operatorname{tr}\log([\varphi_n^{R,*}(e^{i\theta})\varphi_n^{R,*}(e^{i\theta})^\dagger])\\ &= \log\det([\varphi_n^{R,*}(e^{i\theta})\varphi_n^{R,*}(e^{i\theta})^\dagger])\\ &= \log\det([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})])\\ &= \operatorname{tr}\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]).\end{aligned}$$

It remains to apply Lemma 21. $\square$

Corollary 25. *If σ is a Szegő measure, then*

$$\int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi} \leq \inf_n \operatorname{tr}\log\beta_n \leq -\sup_n \log\|\beta_n^{-1}\| \leq 0, \tag{66}$$

in particular, $\sup_n \|\beta_n^{-1}\| < +\infty$.

Proof. Since β_n satisfies (59), all its eigenvalues λ_k, $1 \leq k \leq \ell$, must lie in the interval (0, 1]. In addition,

$$\|\beta_n^{-1}\| = \max_{1\leq k\leq\ell} \lambda_k^{-1}.$$

By (62) and Lemma 3,

$$-\infty < \int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi} \leq \operatorname{tr}\log\beta_n \leq 0,$$

implying that

$$\log\|\beta_n^{-1}\| = \max_k \log\lambda_k^{-1} < \sum_{k=1}^{\ell}\log\lambda_k^{-1} = \operatorname{tr}\log\beta_n^{-1} \leq -\int_{\mathbb{T}} \operatorname{tr}\log\sigma'\frac{d\theta}{2\pi} < +\infty. \quad \square$$

By compactness of closed balls in the finite-dimensional space $\mathcal{M}_\ell$, a bounded sequence of matrices has a limit point. It follows that if $\{\beta_n^{-1}\}_{n\geq0}$ is uniformly bounded, then any of its limit points β^{-1} in $\mathcal{M}_\ell$ satisfies

$$\|\beta^{-1}\| \leq \sup_n \|\beta_n^{-1}\| < +\infty. \tag{67}$$

We denote by $C(\mathbb{T}, \mathcal{M}_\ell^+)$ the set of all continuous matrix functions on $\mathbb{T}$ with values in $\mathcal{M}_\ell^+$, by $L^2(\mathbb{T}, \mathcal{M}_\ell^+)$ the set of all square-integrable matrix functions on $\mathbb{T}$ with values in $\mathcal{M}_\ell^+$, and by $M(\mathbb{T}, \mathcal{M}_\ell^+)$ the set of all finite Borel measures with values in $\mathcal{M}_\ell^+$.

Let μ be a finite Borel measure with values in $\mathcal{M}_\ell^+$. Suppose that, for any open arc $I \subset \mathbb{T}$ whose endpoints do not carry point masses of σ, the inequality

$$\mu(I) \preceq \sigma(I)$$

holds. Then we write $d\mu \preceq d\sigma$.

Theorem 26. *Let $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ satisfy $\sup_n \|\beta_n^{-1}\| < +\infty$, with β_n defined as above, let $\{f_n\}_{n\geq 0}$ be a sequence in $C(\mathbb{T}, \mathcal{M}_\ell^+)$ such that $f_n(e^{i\theta}) \succ \mathbf{0}$ on $\mathbb{T}$, and let*

$$\int_{\mathbb{T}} f_n \frac{d\theta}{2\pi} \preceq \mathbf{1}; \tag{68}$$

$$*-\lim_n f_n \frac{d\theta}{2\pi} \preceq d\sigma; \tag{69}$$

$$\log \beta_n \preceq \int_{\mathbb{T}} \log f_n \frac{d\theta}{2\pi}. \tag{70}$$

Then

$$\lim_n \log \beta_n = \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi} \tag{71}$$

in $\mathcal{M}_\ell$.

Proof. By (59) and (67), the sequence of negative operators $\log \beta_n$ is uniformly bounded. Suppose that $\log \beta$ is a limit point of this sequence of matrices in $\mathcal{M}_\ell$. Then there is an infinite subset Λ of $\mathbb{N}$ such that

$$\lim_{n\in\Lambda} \log \beta_n = \log \beta. \tag{72}$$

Let

$$\log^+ x = \max(\log x, 0), \qquad \log^- x = \log^+ x - \log x.$$

Then $\log^+ x \leq x$ for every $x > 0$. The spectral theorem applied to a (strictly) positive operator A yields

$$\log^+(A) \preceq A. \tag{73}$$

We apply (73) to $A := f_n(e^{i\theta})$ pointwise in θ and obtain the operator inequality

$$\log^+(f_n(e^{i\theta})) \preceq f_n(e^{i\theta}). \tag{74}$$

Integrating (74) and taking into account (68), we obtain

$$\int_{\mathbb{T}} \log^+(f_n(e^{i\theta})) \frac{d\theta}{2\pi} \preceq \mathbf{1}. \tag{75}$$

Observing that $\log = \log^+ - \log^-$ and using (70) and (75), we see that

$$\int_{\mathbb{T}} \log^-(f_n(e^{i\theta})) \frac{d\theta}{2\pi} \preceq \mathbf{1} + \log \beta_n^{-1}. \tag{76}$$

Let

$$d\nu_n^+ := \log^+(f_n(e^{i\theta}))\frac{d\theta}{2\pi}, \qquad d\nu_n^- := \log^-(f_n(e^{i\theta}))\frac{d\theta}{2\pi}. \tag{77}$$

Since $\{\beta_n^{-1}\}_{n\geq 0}$ is bounded, (76) implies that $\{\nu_n^-\}_{n\geq 0}$ has a $*$-weak limit point $\nu^- \in M(\mathbb{T}, \mathcal{M}_\ell^+)$:

$$d\nu^- = *-\lim_{n\in\Lambda'} d\nu_n^- = (\nu^-)'\frac{d\theta}{2\pi} + d\nu_s^- \quad \text{for some } \Lambda' \subset \Lambda, \tag{78}$$

where $d\nu_s^-$ is the singular part of $d\nu^-$ (it may include the discrete part as well), $(\nu^-)' = d\nu^-/(\frac{d\theta}{2\pi})$. It follows from the inequality $(\log^+ x)^2 \leq x$ and (68) that

$$\int_{\mathbb{T}} \left(\log^+(f_n(e^{i\theta}))\right)^2 \frac{d\theta}{2\pi} \preceq \mathbf{1}. \tag{79}$$

By (79), the function $d\nu_n^+/(\frac{d\theta}{2\pi})$ is in the unit ball of $L^2(\mathbb{T}, \mathcal{M}_\ell)$, which is compact in the weak topology of $L^2(\mathbb{T}, \mathcal{M}_\ell)$; see Theorem 8. It follows that any $*$-limit point ω of $\{\nu_n^+\}_{n\geq 0}$ in $M(\mathbb{T}, \mathcal{M}_\ell)$ is absolutely continuous with respect to the Lebesgue measure and, moreover, belongs to $L^2(\mathbb{T}, \mathcal{M}_\ell^+)$. Then there exist a subset $\Lambda'' \subset \Lambda'$ and some ω' in the unit ball of $L^2(\mathbb{T}, \mathcal{M}_\ell^+)$ such that

$$\begin{aligned} d\nu^+ &:= *-\lim_{n\in\Lambda''} d\nu_n^+ = \omega'\frac{d\theta}{2\pi}, \qquad *-\lim_{n\in\Lambda''} d\nu_n^- = d\nu^-, \\ d\nu &:= d\nu^+ - d\nu^- = (\omega' - (\nu^-)')\frac{d\theta}{2\pi} - d\nu_s^-; \end{aligned} \tag{80}$$

see (78). Let I be an open arc on $\mathbb{T}$ such that its endpoints do not carry point masses of $d\nu_s^-$ or $d\sigma_s$. By matrix Jensen's inequality from Proposition 11, we get

$$\frac{1}{|I|}\int_I \log(f_n(e^{i\theta}))\frac{d\theta}{2\pi} \preceq \log\left\{\frac{1}{|I|}\int_I f_n(e^{i\theta})\frac{d\theta}{2\pi}\right\}. \tag{81}$$

Applying Helley's Theorem 6 separately to $\{\nu_n^+\}_{n\in\Lambda''}$ and to $\{\nu_n^-\}_{n\in\Lambda''}$, we obtain

$$\lim_{n\in\Lambda''}\frac{1}{|I|}\int_I \log(f_n(e^{i\theta}))\frac{d\theta}{2\pi} = \frac{\nu(I)}{|I|}. \tag{82}$$

Applying Helley's Theorem 6, we derive from (69) the inequality

$$\lim_n \frac{1}{|I|}\int_I f_n(e^{i\theta})\frac{d\theta}{2\pi} \preceq \frac{\sigma(I)}{|I|}. \tag{83}$$

A substitution of (82) and (83) into (81) results in the inequality

$$\frac{\nu(I)}{|I|} \preceq \log\left(\frac{\sigma(I)}{|I|}\right)$$

(here we use the operator continuity of the logarithm; see Proposition 12). It follows from Lebesgue's theorem on differentiation and the operator continuity of the logarithm that

$$\nu' \preceq \log(\sigma') \tag{84}$$

almost everywhere on $\mathbb{T}$. In view of (70) and (84), we obtain

$$\log\beta + \nu_s^-(\mathbb{T}) \preceq \int_{\mathbb{T}} d\nu + \nu_s^-(\mathbb{T}) = \int_{\mathbb{T}} \nu'\frac{d\theta}{2\pi} \preceq \int_{\mathbb{T}} \log\sigma'\frac{d\theta}{2\pi}. \tag{85}$$

Combining (62) with (85), we see that

$$\int_{\mathbb{T}} \operatorname{tr} \log \sigma' \frac{d\theta}{2\pi} = \operatorname{tr} \log \beta \tag{86}$$

and $\operatorname{tr}\nu_s^-(\mathbb{T}) = 0$, so $\nu_s^- = 0$ by the nonnegativity of the measure ν_s^-. It follows that

$$\log \beta \preceq \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi}.$$

Since the traces of the operators on both sides are equal by (86), we invoke Lemma 2 and conclude that

$$\log \beta = \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi}.$$

Since $\log \beta$ is an arbitrary limit point of $\{\log \beta_n\}_{n \geq 0}$, we obtain (71). □

Theorem 27. *Let* $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ *satisfy* $\sup_n \|\beta_n^{-1}\| < +\infty$. *Then*

$$\lim_n \log \beta_n = \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi}.$$

Proof. Set

$$f_n(e^{i\theta}) = [\varphi_n^{R,*}(e^{i\theta})^\dagger \varphi_n^{R,*}(e^{i\theta})]^{-1}$$

in Theorem 26. Then (68) and (69) follow from Theorem 18. Finally, (70) follows from (58). □

Theorem 28 (*[5, Theorem 18]*)**.** *For any* $\sigma \in \mathsf{P}_\ell(\mathbb{T})$,

$$\log \prod_{k=0}^{\infty} \det(\mathbf{1} - \alpha_k^\dagger \alpha_k) = \int_{\mathbb{T}} \operatorname{tr} \log \sigma' \frac{d\theta}{2\pi}. \tag{87}$$

Proof. Since $\det(\mathbf{1} - \alpha_k^\dagger \alpha_k) < 1$ for all k, the sum of the series

$$\sum_{k=0}^{\infty} \log \det(\mathbf{1} - \alpha_k^\dagger \alpha_k)$$

with negative terms satisfies

$$\sum_{k=0}^{\infty} \log \det(\mathbf{1} - \alpha_k^\dagger \alpha_k) \geq \int_{\mathbb{T}} \operatorname{tr} \log \sigma' \frac{d\theta}{2\pi} \tag{88}$$

by Theorem 24. We have two cases. If the series on the left-hand side of (88) diverges, then σ is not a Szegő measure and both sides of (87) equal $-\infty$. If the series on the left-hand side of (88) converges, then

$$\lim_k \log \det(\mathbf{1} - \alpha_k^\dagger \alpha_k) = \lim_k \operatorname{tr} \log(1 - \alpha_k^\dagger \alpha_k) = 0.$$

Since the spectral norm $\|\cdot\|$ is the largest eigenvalue of a positive self-adjoint matrix, it follows that

$$\lim_k \|\alpha_k^\dagger \alpha_k\| = 0.$$

Since $-x \geq \log(1-x)$ for $0 < x < 1$, we see that

$$-\|\alpha_k^\dagger \alpha_k\| \geq \log(1 - \|\alpha_k^\dagger \alpha_k\|) \geq \operatorname{tr}\log(\mathbf{1} - \alpha_k^\dagger \alpha_k),$$

implying that

$$\sum_{k=0}^{\infty} \|\alpha_k^\dagger \alpha_k\| < +\infty.$$

Lemma 21 implies that the $\|\beta_n^{-1}\|$ are bounded. An application of Theorem 27 now completes the proof. $\square$

Corollary 29 (*[5, Theorem 19]*). *A measure $\sigma \in \mathsf{P}_\ell(\mathbb{T})$ is a Szegö measure if and only if*

$$\sum_{k=0}^{\infty} \|\alpha_k^\dagger \alpha_k\| < +\infty.$$

One direction of this corollary was already proved in Theorem 28; the other direction can be obtained analogously to [5].

Corollary 30. *Let σ be a Szegő measure and let $\{\varphi_n\}_{n\geq 0}$ be the orthogonal polynomials in $L^2(d\sigma)$. Then*

$$*-\lim_n d\mu_n = *-\lim_n \log([\varphi_n^{R,*}(e^{i\theta})^\dagger \varphi_n^{R,*}(e^{i\theta})]^{-1}) \frac{d\theta}{2\pi} = \log(\sigma') \frac{d\theta}{2\pi} \tag{89}$$

in the weak topology of $M(\mathbb{T}, \mathcal{M}_\ell)$.

Proof. Apply the proof of Theorem 26 to the measures

$$d\nu_n^+ := d\mu_n^+ := \log^+([\varphi_n^{R,*}(e^{i\theta})^\dagger \varphi_n^{R,*}(e^{i\theta})]^{-1}) \frac{d\theta}{2\pi},$$
$$d\nu_n^- := d\mu_n^- := \log^-([\varphi_n^{R,*}(e^{i\theta})^\dagger \varphi_n^{R,*}(e^{i\theta})]^{-1}) \frac{d\theta}{2\pi}.$$

Taking (85) into account, we obtain

$$\nu' = \log \sigma' \quad \text{a.e. on } \mathbb{T}. \tag{90}$$

The substitution of (90) into the last formula of (80) results in

$$d\nu = \log \sigma' \frac{d\theta}{2\pi} = (\omega' - (\nu^-)') \frac{d\theta}{2\pi}.$$

Since ω in the proof of Theorem 26 was an arbitrary $*$-limit point of $\{\nu_n^+\}_{n\in\Lambda}$, this implies that $*-\lim_{n\in\Lambda'} d\nu_n^+ = \omega' \frac{d\theta}{2\pi}$. Since ν^- was an arbitrary $*$-limit point of $\{\nu_n^-\}_{n\in\Lambda}$, we conclude that $*-\lim_n d\mu_n = \log \sigma' \frac{d\theta}{2\pi}$. $\square$

9. The Helson–Lowdenslager theorem

Since $\Phi_n^{R,*}$ is left orthogonal to $z\mathbf{1}, \ldots, z^n\mathbf{1}$ (see [4, Lemma 3.2]), it is also left orthogonal to any linear combination p of these matrix functions with the coefficients in $\mathcal{M}_\ell$. Take any such combination p. Then

$$\langle\!\langle \Phi_n^{R,*} - p, \Phi_n^{R,*} - p\rangle\!\rangle_L = \langle\!\langle \Phi_n^{R,*}, \Phi_n^{R,*}\rangle\!\rangle_L + \langle\!\langle p, p\rangle\!\rangle_L - \langle\!\langle \Phi_n^{R,*}, p\rangle\!\rangle_L - \langle\!\langle \Phi_n^{R,*}, p\rangle\!\rangle_L^{\dagger}$$
$$= \langle\!\langle \Phi_n^{R,*}, \Phi_n^{R,*}\rangle\!\rangle_L + \langle\!\langle p, p\rangle\!\rangle_L.$$

Since every polynomial Q satisfying $Q(\mathbf{0}) = \mathbf{1}$ is of the form $Q = \Phi_n^{R,*} - p$, we obtain the matrix inequality

$$\langle\!\langle \Phi_n^{R,*}, \Phi_n^{R,*}\rangle\!\rangle_L \preceq \langle\!\langle Q, Q\rangle\!\rangle_L. \tag{91}$$

These facts are also derived for the real line in [4, Formula (2.10)].

It is therefore natural to call the square root of the positive matrix on the left-hand side of (91) the *left operator distance* from $\mathbf{1}$ to $z\mathcal{P}_{n-1}$. Consequently, the usual distance in the left Hilbert space is equal to

$$\operatorname{dist}_L(\mathbf{1}, z\mathcal{P}_{n-1})^2 = \operatorname{tr}(\langle\!\langle \Phi_n^{R,*}, \Phi_n^{R,*}\rangle\!\rangle_L). \tag{92}$$

One easily verifies (see also [4, Lemma 3.1]) that

$$\langle\!\langle \Phi_n^{R,*}, \Phi_n^{R,*}\rangle\!\rangle_L = \langle\!\langle \Phi_n^{R}, \Phi_n^{R}\rangle\!\rangle_R^{\dagger} = \langle\!\langle \Phi_n^{R}, \Phi_n^{R}\rangle\!\rangle_R = (\kappa_n^R)^{-\dagger}\langle\!\langle \varphi_n^{R}, \varphi_n^{R}\rangle\!\rangle_R(\kappa_n^R)^{-1}$$
$$= (\kappa_n^R)^{-\dagger}\mathbf{1}(\kappa_n^R)^{-1} = (\kappa_n^R)^{-\dagger}(\kappa_n^R)^{-1}. \tag{93}$$

The right-hand side of (93) is Hermitian positive definite, and (32) implies that

$$((\kappa_n^R)^{-\dagger}(\kappa_n^R)^{-1})^{1/2} = \rho_{n-1}^R \cdots \rho_0^R = (\mathbf{1} - \alpha_{n-1}\alpha_{n-1}^{\dagger})^{1/2} \cdots (\mathbf{1} - \alpha_0\alpha_0^{\dagger})^{1/2} \tag{94}$$

is the left matrix distance from $\mathbf{1}$ to $z\mathcal{P}_{n-1}$.

Corollary 31. *The identity polynomial* $\mathbf{1}$ *is in the left closure of the sets of matrix polynomials* $z\mathcal{P}_{n-1}$ *if and only if*

$$\exp \int_{\mathbb{T}} \log \sigma' \frac{d\theta}{2\pi} = \mathbf{0}.$$

The distance formula (94) is useful if the parameters $\{\alpha_k\}_{k\geq 0}$ of σ are known. If this is not the case, one can apply an estimate for $(\kappa_n^R)^{-1}$ from below which was obtained by Helson and Lowdenslager in [9].

Now we are in a position to prove the main result of [9].

Theorem 32 (*[9]*). *For every* $\sigma \in \mathsf{P}_\ell(\mathbb{T})$

$$\exp \int_{\mathbb{T}} \frac{1}{\ell} \operatorname{tr} \log \sigma' \frac{d\theta}{2\pi} = \inf_{A,P} \int_{\mathbb{T}} \frac{1}{\ell} \operatorname{tr}[(A+P)^{\dagger} d\sigma (A+P)], \tag{95}$$

where A runs over all matrices with determinant one, and P over all trigonometric polynomials of the form

$$P(e^{i\theta}) = \sum_{k>0} A_k e^{ik\theta}.$$

Proof. Combining Lemma 5 with formula (94), we get

$$\inf_{A\in\mathcal{A}} \frac{1}{\ell}\operatorname{tr} A((\kappa_n^R)^{-\dagger}(\kappa_n^R)^{-1})A^\dagger = [\det((\kappa_n^R)^{-\dagger}(\kappa_n^R)^{-1})]^{1/\ell}$$
$$= \exp\left\{\frac{1}{\ell}\sum_{j=0}^{n-1}\log\det(\mathbf{1}-\alpha_j\alpha_j^\dagger)\right\} = \exp\left\{\frac{1}{\ell}\operatorname{tr}(\log\beta_n)\right\}.$$

Combining this formula with (92), we obtain

$$\inf_{A\in\mathcal{A},\,P\in z\mathcal{P}_{n-1}} \int_{\mathbb{T}} \frac{1}{\ell}\operatorname{tr}[(A+P)d\sigma(A+P)^\dagger]$$
$$= \inf_{A\in\mathcal{A},\,P\in z\mathcal{P}_{n-1}} \int_{\mathbb{T}} \frac{1}{\ell}\operatorname{tr} A(\mathbf{1}+A^{-1}P)d\sigma(\mathbf{1}+A^{-1}P)^\dagger A^\dagger$$
$$= \inf_{\mathcal{A}} \frac{1}{\ell}\operatorname{tr} A((\kappa_n^R)^{-\dagger}(\kappa_n^R)^{-1})A^\dagger = \exp\left\{\frac{1}{\ell}\int_0^{2\pi} \operatorname{tr}\log([\varphi_n^{R,*}(e^{i\theta})^\dagger\varphi_n^{R,*}(e^{i\theta})]^{-1})\frac{d\theta}{2\pi}\right\}.$$

Passing to the limit, we arrive at (95), initially proved via a different method in [9, Theorem 8]. □

Acknowledgments

The research leading to these results was carried out at Technische Universität Berlin and has received funding from the Alexander von Humboldt Foundation under the Sofja Kovalevskaja Prize Programme and from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 259173.

The authors thank the referees for helpful critique.

References

[1] T. Ando, F. Hiai, Operator log-convex functions and operator means, Mathematische Annalen 350 (3) (2011) 611–630.
[2] A.I. Aptekarev, E.M. Nikishin, The scattering problem for a discrete Sturm–Liouville operator, Mat. Sb. (N.S.) 121(163) (3) (1983) 327–358. (in Russian).
[3] R. Bhatia, Matrix Analysis, Springer, New York, 1996.
[4] D. Damanik, A. Pushnitski, B. Simon, The analytic theory of matrix orthogonal polynomials, Surveys in Approximation Theory 4 (2008) 1–85.
[5] P. Delsarte, Y.V. Genin, Y.G. Kamp, Orthogonal polynomial matrices on the unit circle, IEEE Trans. Circuits and Systems CAS-25 (3) (1978) 149–160.
[6] N. Dunford, J.T. Schwartz, Linear Operators. Part I. General Theory, Wiley, New York, 1988, Reprint of the 1958 original.
[7] D.R. Farenick, F. Zhou, Jensen's inequality relative to matrix-valued measures, J. Math. Anal. Appl. 327 (2007) 919–929.
[8] E.J. Hannan, Multiple Time Series, Wiley, New York, 1970.
[9] H. Helson, D. Lowdenslager, Prediction theory and Fourier series in several variables, Acta Math. 99 (1958) 165–202.
[10] R. Horn, C. Johnson, Matrix Analysis, Cambridge, 1990, Corrected reprint of the 1985 original.
[11] S. Khrushchev, Orthogonal Polynomials and Continued Fractions. From Euler's Point of View, Cambridge, 2008.
[12] S. Khrushchev, Schur's algorithm, orthogonal polynomials, and convergence of Wall's continued fractions in $L^2(\mathbb{T})$, J. of Approx. Theorey 108 (2001) 161–248.
[13] P. Masani, Wiener's contributions to generalized harmonic analysis, prediction theory and filter theory, Bull. Amer. Math. Soc. 72 (1) (1966) 73–125. Part 2.

[14] Yu.A. Rozanov, Stationary Random Processes, Holden-Day, San Francisco, 1967.
[15] B. Simon, Orthogonal Polynomials on the Unit Circle, Part I: Classical Theory, vol. 54, AMS Colloquium Publications, 2005, Part 1.
[16] G. Szegő, Über den asymptotischen Ausdruck von Polynomen, die durch eine Orthogonalitätseigenschaft definiert sind, Math. Ann. 86 (1922) 114–139.
[17] S. Verblunsky, On positive harmonic functions, Proc. London Math. Soc. 38 (2) (1935) 125–157.
[18] N. Wiener, P. Masani, The prediction theory of multivariate stochastic processes, I. The regularity conditions, Acta Math. 98 (1957) 111–150.